BRIDGING THE SEMANTIC GAP USING RANKING SVM FOR IMAGE RETRIEVAL

Haiying Guan, Sameer Antani, L. Rodney Long, and George R. Thoma

Lister Hill National Center for Biomedical Communications
National Library of Medicine, National Institutes of Health
guanh@mail.nih.gov

ABSTRACT

One of the main challenges for Content-Based Image Retrieval (CBIR) is to achieve meaningful mappings between the high-level semantic concepts and the low-level visual features in images. This paper presents an approach for bridging this semantic gap to improve retrieval quality using the Ranking Support Vector Machine (Ranking SVM) algorithm. Ranking SVM is a supervised learning algorithm which models the relationship between semantic concepts and image features, and performs retrieval at the semantic level. We apply it to the problem of vertebra shape retrieval on a digitized spine x-ray image collection from the second National Health and Nutrition Examination Survey (NHANES II). The experimental results show that the retrieval precision is improved $2.45 - 15.16\%$ using the proposed approach.

***Index Terms*—** Content-Based Image Retrieval, NHANES II database, Ranking SVM, digital radiography

1. INTRODUCTION

In recent years, there has been a significant increase in the use of images in clinical medicine and biomedical research. This has underscored a compelling need for efficient image searching and retrieval techniques. Content-Based Image Retrieval (CBIR) has been discussed in the literature as a possible solution for the problem. Previous research in CBIR has mainly focused on extracting low-level visual features (e.g., color, texture, shape, spatial layout) and then using them directly to compute image similarity. Extensive experiments have shown, however, that low-level image features cannot always capture the semantic concepts in the image [1]. This poses a serious shortcoming in applying CBIR to routine clinical use, where image similarity is defined in terms of biomedical concepts. In general, there is no direct link between such high-level semantic concepts and the automatically extracted, low-level image features. Therefore, to support query by semantic concept, there is a compelling need for CBIR systems to provide maximum support towards bridging the ‘semantic gap’ between the low-level visual features and the semantics in biomedical concepts.

The Relevance Feedback (RF) algorithm [2] combines image browsing with online learning to reduce the semantic gap. It has been shown to be a viable approach in some contexts by using user feedback to iteratively improve retrieval quality through methods such as feature fusion by linear weighting [3]. The method is seriously challenged when, in some cases, it may be inconvenient, unsuitable, or impossible to obtain the user feedback. Additionally, the performance of RF also greatly depends on the retrieval performance in the first round of the retrieval. Other CBIR retrieval algorithms are based on classification algorithms [4] [5]. However, in many applications, even the images that are in the same category can be dissimilar in appearance, e.g., chest x-rays in different views. Another problem for traditional learning-based algorithms such as Support Vector Machine (SVM) [4] and Bayesian classifier [5] is classifying instead of ranking the retrieved images.

The paper presents a retrieval method that attempts to bridge the gap between the low-level feature space and concept-based semantic pathology categories using machine learning techniques. We address the shortcoming in RF algorithms and improve the retrieval performance in the first round of the retrieval without user interaction through off-line learning-based feature fusion. We describe a learning-based approach, the Ranking SVM, for CBIR of medical images. The ranking SVM algorithm is proposed by Herbrich et al. in 2000 [6]. Joachims et al. applied it in the search engine optimization for the text retrieval [7]. We introduce it to the CBIR domain for the image retrieval. The method predicts the ranking function by attempting to bridge the gap between the image pathology (expert-marked ground truth data) and the low-level image features. The retrieval function is learned by optimizing a set of inequalities using SVM techniques to improve retrieval effectiveness and efficiency. We demonstrate our approach for retrieval of vertebral shapes segmented from digitized spine x-ray images from health survey data archived by the National Library of Medicine. Our experiment shows that the proposed approach significantly improves retrieval recall and precision.

2. METHODS

In this section, we introduce a learning approach, the Ranking SVM, that addresses the limitations of standard SVM and

Bayes classifiers, when applied to image retrieval, namely: (i) how to learn the ranking functions instead of classification functions; and (ii) how to bridge the semantic gap.

2.1. The Ranking SVM algorithm

The Ranking SVM algorithm was originally proposed in [7] for search engine optimization for document retrieval. Beginning with the SVM approach, the Ranking SVM uses a method for learning the retrieval function by optimizing a set of inequalities. However, it is different from standard SVM in that it can effectively adapt the retrieval function to a given partial ranking order instead of the document categories of the training samples. Thus, the training samples are now based on the experts' ranking information on the training dataset instead of simple category information. In this paper, we introduce Ranking SVM into the CBIR domain.

We formulate the ranking problem in image retrieval as follows: For a given image query q and a image database collection $D = d_1, ..., d_m$, the optimal retrieval system should return a optimal ordering (ranking) r^* that orders the database elements in D according to their relevance to the query. The objective of the retrieval algorithm is to find a retrieval function f, whose ordering $r_{f(q)}$ approximates the optimum ordering r^*, according to a target standard of optimality.

Typically, retrieval systems do not achieve an optimal ordering r^*. Instead, a retrieval function f is evaluated by how closely its ordering $r_{f(q)}$ approximates the optimum, given an independently and identically distributed training sample S of size n containing queries q with their target rankings r^*, $(q_1, r_1^*), (q_2, r_2^*), ..., (q_n, r_n^*)$. The retrieval problem may be posed as a maximization problem in terms of Kendall's rank correlation coefficient τ [8]; In this formulation, the learner L selects a ranking function f from a family of ranking functions F that maximizes the empirical τ on the training sample. More precisely, the objective is learn a scoring function for ranking; that is, learn to accurately rank a set of objects by combining a given collection of ranking or preference functions.

The input to the algorithm is a list of "order preferences" (e.g., d_i should be ranked above d_j), and a list of "base ranking functions" from a family of ranking functions F. The goal is to find a function $f \in F$ that maximizes Kendall's τ and generalizes beyond the training data. Consider the class of linear ranking functions,

$$(d_i, d_j) \in f_{\overrightarrow{w}}(q) \Longleftrightarrow \overrightarrow{w}^T \Phi(q, d_i) > \overrightarrow{w}^T \Phi(q, d_j), \quad (1)$$

where, $\overrightarrow{w}$ is a $(p \times 1)$ weight vector that is adjusted by learning. $\Phi(q, d)$ is a mapping, $\Phi : Q \times D \to R^p$, onto features that describe the match between query q and the image d, where $Q = possible\ query\ images$, and $D = database\ images$. A possible formulation of the retrieval problem is finding the weight vector that minimizes the following equation,

$$V(\overrightarrow{w}, \overrightarrow{\xi}) = \frac{1}{2} \overrightarrow{w}^T \cdot \overrightarrow{w} + C \sum \xi_{i,j,k}, \quad (2)$$

subject to the constraints of the inequalities:

$$\forall (d_i, d_j) \in r_1^* : \overrightarrow{w}^T \Phi(q_1, d_i) \geq \overrightarrow{w}^T \Phi(q_1, d_j) + 1 - \xi_{i,j,1} \quad (3)$$
$$...$$
$$\forall (d_i, d_j) \in r_n^* : \overrightarrow{w}^T \Phi(q_n, d_i) \geq \overrightarrow{w}^T \Phi(q_n, d_j) + 1 - \xi_{i,j,n} \quad (4)$$
$$\forall i \forall i \forall k : \xi_{i,j,k} \geq 0.$$

C is a parameter that allows trading-off margin size against training error. ξ is introduced as a non-negative slack variable, such that the upper bound of $\sum \xi_{i,j,k}$ is minimized. The optimization problem is equivalent to that of a classification SVM on pairwise difference vectors $\Phi(q_k, d_i) - \Phi(q_k, d_j)$. It can be solved using decomposition algorithms similar to those used for SVM classification. An adaptation of the $SVM^{light} algorithm$ [9] [1] is used in the experiment.

An advantage of the Ranking SVM algorithm is that it is theoretically well-founded: Kendall's τ gives a precise ranking order measurement and it is directly related to Average Precision (i.e. average of precision at each relevant item retrieved). Another advantage of the Ranking SVM is that it inherits the SVM algorithm's property of the maximum-margin approach [9], which can avoid the "curse of dimensionality" for classification even without a feature-selection step that is essential for many conventional methods. In addition, the method is a constructive learning procedure based on statistical learning theory. It is based on the principle of structural risk minimization, which aims at minimizing the bound on the generalization error (i.e., error made by the learning machine on data unseen during training) rather than minimizing the mean square error over the data set. As a result, Ranking SVM tends to perform well when applied to data outside the training set.

3. EXPERIMENT AND EVALUATION

We apply the Ranking SVM approach to vertebra shape retrieval and demonstrate that it outperforms a low-level image feature based similarity measure such as the Minkowski metric [10]. The Ranking SVM method effectively models high-level semantic concepts and is able to search for semantically meaningful similar shapes.

3.1. Data set and ground-truth

A collection of 200 cervical spine (C-spine) x-ray images were arbitrarily selected from a collection of 17,000 digitized spine x-ray image from the second National Health and Nutrition Examination Survey (NHANES II) that is archived by the National Library of Medicine [11]. Osteophytes are among various pathologies that are reliably and frequently detectable in the collection. These abnormal vertebrae are identified by the subtle abnormal shape variation due to the anterior osteophytes (AO) as seen on the anterior corner of the vertebra

[1] http://svmlight.joachims.org

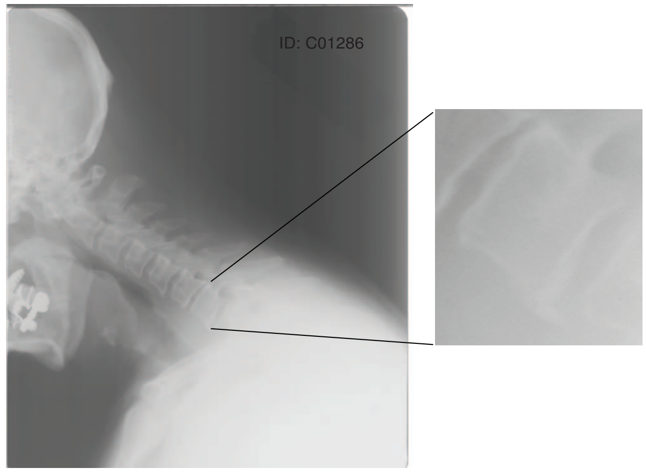

Fig. 1. C-spine x-ray and C_7 with severe claw AO

outline in the sagittal view. Each image contains 6-8 vertebra shapes that have been labeled by three board-certified radiologists according to two classification schemes: (i) Macnab's classification [12], which defines types of osteophytes (claw and traction); and (ii) a grading system for assigning severity levels to AO [13] (slight, moderate, and severe). This results in ten classes: normal, slight claw, slight traction, slight claw and traction, moderate claw, moderate traction, moderate claw and traction, severe claw, severe traction, severe claw and traction. An example C-spine image and a C_7 vertebra with claw type osteophyte of severe grade is shown in Figure 1. Due to the observer biases and variabilities of the medical experts, the AO type and grade labeling is inconsistent to some degree, especially in cases that are borderline between the definitions. To ameliorate this difficulty, for this experiment, we selected 117 vertebrae with consensus on multiple expert opinions on the anterior inferior corner. The method was trained and tested on this data set using the leave-one-out algorithm.

3.2. Image preprocessing and feature extraction

The vertebra shapes were segmented using an Active Contour Segmentation algorithm [14]. After curve fitting, smoothing, and re-sampling, each vertebra was represented with 180 boundary points. Finally, the whole vertebra shape and the AO corner shape features are extracted and normalized using Min-Max normalization. The global shape features are geometric (elongation, eccentricity, roughness, and compactness), Fourier descriptors with complex coordinates [15], Fourier descriptors with Centroid Contour Distance Curve [15], Fourier Coefficient of Fourier Expansion of Bent function [16], and moment invariants. The local shape features are turn angle and Distance Across the Shape [17].

3.3. Evaluation

The training data set of the Ranking SVM was ranked according to the AO type and the grade, as described. The coarse ranking order of the training samples was determined by the AO class, i.e. the shapes in the data set that are in the same class as a query were ranked higher (similar, smaller distance) with respect to the query than those in another class. The strict order among the training samples within the same class are determined by the ground truth ranking, (e.g., non-metric similarity measurement such as Procrustes distance, or expert assigned ranking).

The proposed method was trained and tested on the data set using the leave-one-out procedure. Specifically, a vertebra shape was iteratively selected as the test set and the remainder were used for Ranking SVM training. The test results were then averaged over all runs to compute overall performance.

In order to provide an objective comparison of the retrieval performance of the algorithms, we used the quantitative evaluation criterion, the average precision-recall graph. Retrieval precision is defined as the proportion of the images among all those retrieved that are truly relevant to a given query; recall is defined as the proportion of the images that are actually retrieved among all the relevant images to a query. The average precision-recall graph is a plot of the average retrieval precision vs. the average recall over the precision and recall operating ranges of interest. We considered an image to be truly relevant to a query if the retrieved images were in the same class (both have the same type and grade) as the query image. This is a very strict criterion for the retrieval quality evaluation, because the ground truth of the vertebra corner type and grade may even confound the medical experts due to the subtle variations of the corner shapes. The retrieval

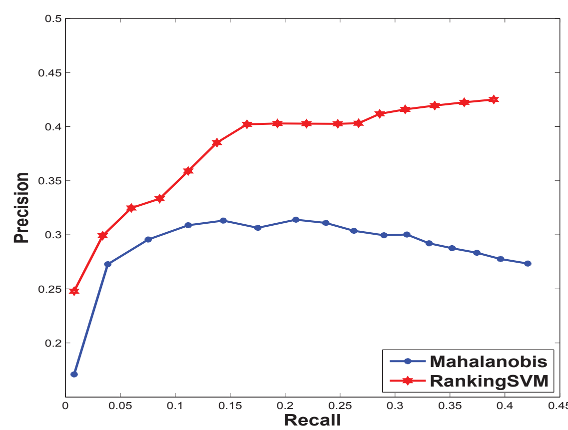

Fig. 2. The Average Precision Recall graph for inferior corner C-spine data set

quality of the Ranking SVM algorithm was compared with the conventional Minkowski metric-based similarity measure, which is Mahalanobis distance. Mahalanobis differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant. Specifically, we performed retrieval for each query image in two different ways: (1) by using Ranking SVM and (2) by calculating image similarities to the query image using Mahalanobis distance. The comparison of the average precision recall graph is shown in Fig. 2. The result also shows that the average retrieval precision of the Ranking SVM is about $2.45 - 15.16\%$ better than the conventional similarity measurement. Unlike the existing approaches to CBIR, which are typically based on some simple distance measures for image similarity, the Ranking SVM is an off-line learning based approach, which is trained to predict the measures of image similarity that are implicit in expert-labeled training data. It treats the learning of the similarity function as an optimization problem and seeks to effectively learn how similarity ranking using high-level human perception may be mapped to ranking using low-level image features. In our work, this learned similarity measurement (ranking function) performs much better than the conventional Mahalanobis distance similarity measurement.

4. CONCLUSION

In this paper, we have introduced the Ranking SVM, a machine learning-based algorithm, for modeling semantic concepts using low-level image features for CBIR. We have described the underlying theory and evaluated the method for retrieval of vertebra shapes with osteophytes of varying severity from C-spine x-ray images. The results demonstrate that the learning algorithm may be able to approximate the ranking function implicit in the expert marked and labeled image data set. The training data set encodes pathological semantic concepts, and serves as a basis for retrieving visually similar vertebrae. Furthermore, the retrieved vertebra shapes may have predictive value for the disease condition of the query. It also shows that the precision of the learning-based framework significantly outperforms the commonly used simple distance-based similarity metric by $2.45 - 15.16\%$ in the experiment.

5. ACKNOWLEDGEMENT

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

6. REFERENCES

[1] Y. Liu, D. Zhang, G. Lu, and W.Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.

[2] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, September 1998.

[3] X. Xu, D.J. Lee, S. Antani, and L.R. Long, "Relevance feedback for spine x-ray retrieval," in *Proc. of the 18th International Symposium on Computer-Based Medical Systems*, 2005, pp. 197–202.

[4] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *MULTIMEDIA '01: Proceedings of the Ninth ACM International Conference on Multimedia*, New York, NY, USA, 2001, pp. 107–118, ACM.

[5] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, and P.N. Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments," *IEEE Transactions on Image Processing*, vol. 9, pp. 20–37, 2000.

[6] Ralf Herbrich, Thore Graepel, and Klaus Obermayer, *Large Margin Rank Boundaries for Ordinal Regression*, MIT Press, Cambridge, MA, 2000.

[7] T. Joachims, "Optimizing search engines using clickthrough data," in *KDD '02: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2002, pp. 133–142, ACM Press.

[8] A.M. Mood, F.A. Graybill, and D.C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill Companies, 3 edition, 1974.

[9] T. Joachims, *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*, Kluwer Academic Publishers, April 2002.

[10] Gregory L. Naber, *The Geometry of Minkowski Spacetime: An Introduction to the Mathematics of the Special Theory of Relativity*, Dover Publications, 2003.

[11] D.J. Lee D.M. Krainak G.R. Thoma L.R. Long, S. Antani, "Biomedical information from a national collection of spine x-rays: Film to content-based retrieval," in *Proc. of SPIE Medical Imaging: PACS and Integrated Medical Information Systems*, February 2003, pp. 5033:70–84.

[12] M.H. Heggeness and B.J. Doherty, "Morphologic study of lumbar vertebral osteophytes," *Southern Medical Journal*, vol. 91, no. 2, pp. 187–189, February 1998.

[13] J. Shou, S. Antani, L.R. Long, and G. R. Thoma, "Evaluating partial shape queries for pathology-based retrieval of verterbra," in *Proc. 8th World Multiconference on Systemics, Cybernetics and Informatics*, July 2004, pp. 155–60.

[14] L.R. Long, S. Antani, and G. Thoma., "Image informatics at a national research center," *Computerized Medical Imaging and Graphics*, vol. 29, pp. 171–193, 2005.

[15] D. Zhang and G. Lu, "Content-based shape retrieval using different shape descriptors: A comparative study," *IEEE International Conference on Multimedia and Expo*, vol. 00, pp. 1139–1142, 2001.

[16] D.-J. Lee, P. Zhan, A. Thomas, and R. B. Schoenberger, "Shape-based human detection for threat assessment," 2004, vol. 5438, pp. 81–91, SPIE.

[17] L. Chen, R. Feris, and M. Turk, "Efficient partial shape matching using Smith-Waterman algorithm," in *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, Anchorage, Alaska, 2008, pp. 23–28.